



TITLE:

Parameter Estimation of Markov Random Fields (時系列における統計的推定論の研究)

AUTHOR(S):

KUNSCH, H.

CITATION:

KUNSCH, H.. Parameter Estimation of Markov Random Fields (時系列における統計的推定論の研究). 数理解析研究所講究録 1977, 312: 173-184

ISSUE DATE:

1977-11

URL:

<http://hdl.handle.net/2433/103906>

RIGHT:

PARAMETER ESTIMATION OF MARKOV RANDOM FIELDS .

東大 理学部 H. Künsch

0. Introduction: A random field is a stochastic process with a multidimensional parameter, usually interpreted as space. There are many examples of random fields in practical applications, for instance annual rainfall in an area, the size of a crop at different plots, or in the case of transmission of photos the optical amplitude. The paper [6] contains a survey of applications of random fields in various disciplines.

We consider here the case of a random field $X(t)$, $t \in \mathbb{Z}^v$, $v = 1, 2, \dots$ with $EX(t) \equiv 0$ and which is homogeneous, i.e. all finite dimensional distributions are translation inva-

variant (A generalization to a field with some trend is possible). Whittle [7] and Larimore [2] both considered autoregressive models, which seem to us a too narrow class; we take here the broader class of Markov models. Two methods for the estimation of the parameters of such a model are proposed, and in both cases consistency and asymptotic normality are proved, even if X is not Gaussian.

1. Type of models; a) Markovian models

Choose a finite set $L \subset \mathbb{Z}^n$ with $L = -L$ and $0 \notin L$, the set of 'points near 0'. We define then the (L -) boundary of a $D \subset \mathbb{Z}^n$ by $\partial D := \{t \notin D \mid \exists s \in D, t-s \in L\}$.

Definition: A random field $X(t), t \in \mathbb{Z}^n$, has the Markov property w.r. to a set D if

$\forall \sigma\{X(t), t \in D\}$ - measurable r. v. U

$$E(U \mid X(t), t \notin D) = E(U \mid X(t), t \in \partial D) \quad (1)$$

Remark: If $n=1$ and $L = \{-p, \dots, -1, 1, \dots, p\}$, then for $D = \{t, t+1, \dots\}$ the above definition is the usual definition of a p -th order Markov

process. It can also be proved, that a p -th order Markov process has the Markov property w.r. to all $D \subset \mathbb{Z}'$. However, for $v \geq 2$ the line transect $X(t, 0, \dots, 0)$ of a Markov field is in general not a Markov process.

If X is homogeneous, Gaussian, $EX = 0$ then the Markov property w.r. to $\{t\}$ is simply $E(X(t) | X(s), s \neq t) = \sum_{k \in L} a(k) X(t-k)$ (2) with some coefficients $a(k)$. (2) is equivalent to $X(t) = \sum_{k \in L} a(k) X(t-k) + Y(t)$ (3) with $Y(t)$ independent of $X(s), s \neq t$. As our models, we take solutions of (2), resp. (3).

A simple calculation shows, that

$$EY(t)Y(s) = c^2 \begin{cases} 1 & \text{if } t=s \\ a(t-s) & \text{if } t-s \in L \\ 0 & \text{elsewhere.} \end{cases}$$

i.e. the error terms in (3) are correlated between themselves. Also, we must have $a(k) = a(-k)$ (therefore $L = -L$).

Theorem 1: There exists a Gaussian homogeneous solution of (2) with $EY(t)^2 = c^2 > 0$ iff $a(k) = a(-k)$, $P(x) := 1 - \sum a(k)e^{ikx} \geq 0$, $\int_{[\pi, \pi]^v} \frac{1}{P(x)} dx < \infty$

If we ask X to be purely non deterministic, then X is uniquely given by

$$E X(t) X(s) = R(t-s) := \frac{c^2}{(2\pi)^\nu} \int \frac{e^{i(t-s)x}}{P(x)} dx \quad (4)$$

Proof: see Rozanov [5]

X can be built up with i.i.d. r.v.:

Theorem 2: If the covariance of X is given by (4), ^{then} X satisfies the moving average equation

$$X(t) = \sum_{k \in \mathbb{Z}^\nu} g(t-k) U(k), \text{ with} \quad (5)$$

$$E U(t) U(s) = c^2 \delta_{t,s} \quad g(s) = \frac{1}{(2\pi)^\nu} \int e^{isx} P(x)^{-\frac{1}{2}} dx$$

Proof: Put $U(k) := \int e^{ikx} P(x)^{\frac{1}{2}} dZ(x)$, where dZ is the random measure with $E |dZ(x)|^2 = \frac{c^2}{(2\pi)^\nu} P(x)^{-1} dx$.
q.e.d.

If U is not normal, the field X defined by (5) is in general only 'Markov in the weak sense'

$$\text{i.e. } \hat{E}(X(t) | X(s), s \neq t) = \sum_{k \in L} a(k) X(t-k)$$

where \hat{E} is the orthogonal projection on the closed linear hull. But since it is not known, if (3) has a non Gaussian solution, we take nevertheless X defined by (5) as our model.

If $X(t)$ has been observed for $t \in D$ and

if $X(t)$ satisfies (2) with known $a(k)$, then the formula for best linear extrapolation of $X(t)$, $t \notin D$, is well known: If D^c is finite, see Rozanov [4], chapter II. 10, if D is a halfspace, see Pitt [3], proposition 9.1, and if D is finite and $\sum |a(k)| < 1$, see Williams [8]

b) Autoregressive models

Definition: X is called an autoregressive field if
$$X(t) = \sum_{k \in K} b(k) X(t-k) + U(t) \quad (6)$$
 with i.i.d. $U(t)$ and K finite, $\neq 0$.

From part 1a) it is clear, that then $U(t)$ must be correlated with some $X(s)$, $s \neq t$.

Theorem 3: There exists a purely non deterministic solution of (6) iff

$$Q(x) := \left[1 - \sum_{k \in K} b(k) e^{-ikx} \right]^{-1} \in L_2([- \pi, \pi]^0, dx).$$

The covariance is $EX(t)X(s) = c^2 \int \frac{e^{i(t-s)x}}{|Q(x)|^2} dx$

Proof: Purely non deterministic implies, that the spectral measure is abs. continuous (Rozanov [5]). Then $EU(t)U(s) = \int e^{i(t-s)x} |Q(x)|^2 f(x) dx$. So, since the U 's are i.i.d. $|Q(x)|^2 f(x) = \text{const.}$
q. e. d.

Therefore every autoregressive field is Markov, and the contrary is true, iff $P(x) = \text{const.} |Q(x)|^2$. For $v=1$, every $P(x)$ has such a decomposition, (see Yaglom [9], p. 121), but for $v \geq 2$, this is no longer the case; for instance the simplest $P(x) = 1 - a \cos x_1 - b \cos x_2$, $a \neq 0, b \neq 0$ has no such decomposition: If we take $Q(x) = 1 - \alpha e^{ix_1} - \beta e^{ix_2} - \gamma e^{-ix_1} - \delta e^{-ix_2}$, then in $|Q|^2$ appear terms of order 2, which are not all $= 0$ except if $\beta = \gamma = \delta = 0$. For more complicated Q 's a similar thing is true.

2. Estimation of parameters

Suppose $X(t)$ is observed in a set D . We want to estimate the $a(k)$'s with the help of $X(t)$, $t \in D$. L is supposed to be known. Since we must have $a(k) = a(-k)$, we choose an M such that $M \cap (-M) = \emptyset$, $M \cup (-M) = L$ and estimate $a(k)$, $k \in M$. For simplicity, we take $D = \{1, \dots, T\}^v$.

The properties of the estimators, we go to propose ~~in the~~^{later} following, are based

on the properties of the following r.v.

$$C(k) := \frac{1}{T^{\nu}} \sum_{t \in D} X(t) X(t+k) \quad (X(s) := 0 \quad \forall s \notin D)$$

Lemma 1: If X is given by (5) with $EU(t)^4 = 3c^4 + \kappa_4 < \infty$, then $C(k) \xrightarrow[T \rightarrow \infty]{P} R(k)$. If moreover $P(x) \neq 0 \quad \forall x$, then $T^{\nu/2}[C(k) - R(k)]$ are asymptotically joint normal with mean 0 and covariance $c^4 A_{k,e} + \frac{\kappa_4}{c^4} R(k)R(e)$, where $A_{k,e} := \frac{2}{(2\pi)^{\nu}} \int \frac{\cos kx \cos ex}{P(x)^2} dx$

Proof: Consistency follows easily, because the field $X(t)X(t+k)$, $t \in \mathbb{Z}^{\nu}$ has a spectral density. Asymptotic normality is a straightforward generalization of Anderson [1], section 8.4.2.

q.e.d.

a) Least square estimators: We take as estimators for $a(k)$ those values $\hat{a}(k)$, which minimize $\sum_{t \in D} [X(t) - \sum_{k \in M} a(k) \{X(t+k) + X(t-k)\}]^2$. By differentiation and division by $4 \cdot T^{\nu}$:

$$\sum_{k \in M} [C(k+n) + C(k-n)] \hat{a}(k) = C(n) \quad n \in M \quad (7)$$

As estimation for c^2 we take

$$\hat{c}^2 := \frac{1}{T^{\nu}} \sum_{t \in D} [X(t) - \sum_{k \in M} \hat{a}(k) \{X(t+k) + X(t-k)\}]^2 \quad (8)$$

$$\text{or by (7)} \quad \hat{c}^2 = C(0) - 2 \sum_{k \in M} \hat{a}(k) C(k) \quad (8')$$

Put $S_{k,n} := R(k+n) + R(k-n)$. Then

Lemma 2: $(S_{k,n})_{k,n \in M}$ is not singular

Proof: $\sum_n S_{k,n} x_n = 0 \quad \forall k \in M \Leftrightarrow$

$$\int e^{ikx} \sum_n x_n \cos nx \frac{dx}{P(x)} = 0 \quad \forall k \in L \Leftrightarrow \sum_n x_n \cos nx = 0$$

$$\text{in } L_2\left(\frac{dx}{P(x)}\right) \Leftrightarrow x_n = 0 \quad \forall n \quad \text{q.e.d.}$$

Theorem 4: If X satisfies (5) with $EU(t) < \infty$ then $\hat{a}(k), \hat{c}^2$ are consistent. If moreover $P(x) \neq 0 \quad \forall x$, then $T^{1/2}[\hat{a}(k) - a(k)]$ are asymptotically joint normal with mean 0 and covariance $c^4 \cdot (S_{k,n})^{-2}$

Proof (3) implies $R(n) - \sum_{k \in M} a(k) S_{k,n} = c^2 \delta_{n,0}$ so consistency follows from lemmas 1 and 2. For asymptotic normality, observe that

$$T^{1/2}[\hat{a}(k) - a(k)] \sim \sum_n (S^{-1})_{k,n} T^{1/2} \left[C(n) - \sum_{\ell} a(\ell) \{C(\ell+n) + C(\ell-n)\} \right]$$

$$\sim \sum_n (S^{-1})_{k,n} T^{1/2} \left[C(n) - R(n) - \sum_{\ell} a(\ell) \{C(\ell+n) - R(\ell+n) + C(\ell-n) - R(\ell-n)\} \right],$$

and the result follows from lemma 1 by an easy calculation q.e.d.

b) Maximum likelihood estimators

Using a result from Whittle [7], p. 440-442, we get that

$$\frac{1}{T^\nu} \log \text{likelihood} \approx \text{const.} - \frac{1}{2} \log c^2 + \frac{1}{2} \frac{1}{(2\pi)^\nu} \int \log(1 - 2 \sum_k a(k) \cos kx) dx$$

$$- \frac{1}{c^2} \left[C(0) - 2 \sum_k a(k) C(k) \right] \quad \dots$$

This gives after differentiation the following equations for the maximum likelihood estimators:

$$\frac{\tilde{C}^2}{(2\pi)^\nu} \int \frac{\cos kx}{1 - 2 \sum_l \tilde{a}(l) \cos lx} dx = C(k), \quad k \in M \quad (9)$$

$$\tilde{C}^2 = C(0) - 2 \sum_k \tilde{a}(k) C(k), \quad (10)$$

Or equivalently to (9) and (10):

$$\frac{\tilde{C}^2}{(2\pi)^\nu} \int \frac{\cos kx}{1 - 2 \sum_l \tilde{a}(l) \cos lx} dx = C(k), \quad k \in M \cup \{0\} \quad (11)$$

i.e. we have to fit in $\tilde{C}^2, \tilde{a}(k)$ in such a way, that for $k \in M \cup \{0\}$ the covariances of the fitted model are equal to $C(k)$. The least square method on the other hand uses also $C(l), l \notin M \cup \{0\}$ for the estimation.

Theorem 5 If X satisfies (5) with $EU(t)^4 < \infty$ and $P(x) \neq 0 \forall x$, then $\tilde{a}(k)$ and \tilde{C}^2 are consistent and $T^{\nu/2} [\tilde{a}(k) - a(k)]$ are asymptotically joint normal with mean 0 and covariances $(A_{k,l} - 2 \frac{R(k)R(l)}{C^4})^{-1}$ ($A_{k,l}$ as in lemma 1).

Proof: $\varphi: (a_k)_{k \in M}, C^2 \longrightarrow \frac{C^2}{(2\pi)^\nu} \int \frac{\cos kx}{1 - 2 \sum_l a_l \cos lx} dx$,
 $k \in M \cup \{0\}$ is a differentiable map defined on $\{a_k \mid 1 - \sum_l a_l \cos lx > 0 \forall x\} \times \mathbb{R}_+$ with functional matrix

$$\left\{ \underbrace{c^2 A_{k,e}}_{c \in M} \left| \frac{R_k}{c^2} \right. \right\}_{k \in M \cup \{0\}}$$

By a similar argument as in lemma 2, we can show, that it is not singular. Together with lemma 1, this implies consistency.

if we develop (9) in a Taylor series, we get: $C(k) - R(k) = \sum_e [\tilde{a}(e) - a(e)] c^2 A_{k,e} + (\tilde{c}^2 - c^2) \frac{R(k)}{c^2} + \text{remainder}$, $k \in M$

Out of (10): $\tilde{c}^2 - c^2 = C(0) - R(0) - 2 \sum_e R(e) [\tilde{a}(e) - a(e)] - 2 \sum_e a(e) [C(e) - R(e)] + \text{remainder}$

Combining these two results:

$$T^{1/2} [C(k) - R(k) - \frac{R(k)}{c^2} \cdot Z] = \sum_e (c^2 A_{k,e} - 2 \frac{R(k)R(e)}{c^2}) T^{1/2} [\tilde{a}(e) - a(e)] + \text{remainder}, \text{ where } Z := C(0) - R(0) - 2 \sum_e a(e) [C(e) - R(e)].$$

The remainder goes to 0 as $T \rightarrow \infty$, and by lemma 1, the left hand side is asymptotically joint normal with covariance $c^4 A_{k,e} - 2 R(k)R(e)$. As in lemma 2, we can show, that this matrix is not singular. g.e.d

c) Comparison of the 2 methods In the case $v=1$, $L = \{-1, 1\}$, we get for the asymptotic variance of $T^{1/2}(\hat{a} - a)$: $\frac{(1-\beta^2)^2}{(1+\beta^2)^4}$ and for

$T^{1/2}(\tilde{a}-a) : \frac{(1-\beta^2)^3}{(1+\beta^2)^4}$, where $a = \frac{\beta}{1+\beta^2}$. So, for $\beta = 0.5 \Leftrightarrow a = 0.4$, we need in the case of the least square method $\frac{1}{3}$ more observations for the same precision, but equation (7) is easier to solve than (11). We haven't had time and occasion to solve numerical examples. We think, it might be a good method to calculate first $\hat{a}(k)$, \hat{c}^2 and use these values as starting values for the solution of (11) with Newton's method.

References

- [1] T.W. Anderson, The statistical analysis of time series, Wiley 1971
- [2] W.E. Larimore, Statistical inference on stationary random fields, Proceedings IEEE 65 (1977), 961-70
- [3] L.D. Pitt, A Markov property for Gaussian processes with a multidimensional parameter, Archive rational mechanics analysis 43 (1971), 343-65
- [4] Yu. A. Rozanov, Stationary random processes, Holden Day 1967

- [5] Yu. A. Rozanov, On Gaussian fields with given conditional distributions, Theory Prob. Appl 12 (1967), 381 - 391
- [6] D. Unwin - L. Hepple, The statistical analysis of spatial series, The statistician 23 (1974), 211-27
- [7] P. Whittle, On stationary processes in the plane, Biometrika 41 (1954), 434 - 449
- [8] D. Williams, Basic theorems on Harnesses, in Stochastic Analysis, D. Kendall - E. Harding ed., Wiley 1973, 349 - 363
- [9] A. M. Yaglom, Stationary random functions, Prentice - Hall 1962.

Present address:

H. Künsch

Hügelstraße 11

8002 Zürich Switzerland.